UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/559,386	05/24/2006	Timo Surakka	108800-00007	4809

4372   7590   12/10/2008
ARENT FOX LLP
1050 CONNECTICUT AVENUE, N.W.
SUITE 400
WASHINGTON, DC 20036

EXAMINER
HOANG, SON T

ART UNIT	PAPER NUMBER
2165	

NOTIFICATION DATE	DELIVERY MODE
12/10/2008	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

DCIPDocket@arentfox.com
IPMatters@arentfox.com
Patent_Mail@arentfox.com

	Application No.	Applicant(s)
Office Action Summary	10/559,386	SURAKKA ET AL.
	Examiner	Art Unit
	SON T. HOANG	2165

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>03</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>14 November 2008</u>.
2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.
3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-31* is/are pending in the application.
   4a) Of the above claim(s) _____ is/are withdrawn from consideration.
5)☐ Claim(s) _____ is/are allowed.
6)☒ Claim(s) *1-31* is/are rejected.
7)☐ Claim(s) _____ is/are objected to.
8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.
10)☒ The drawing(s) filed on <u>06 December 2005</u> is/are: a)☒ accepted or b)☐ objected to by the Examiner.
   Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
   Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
   a)☒ All   b)☐ Some * c)☐ None of:
    1.☒ Certified copies of the priority documents have been received.
    2.☐ Certified copies of the priority documents have been received in Application No. _____.
    3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
   * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
      Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
      Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1.   A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on November 14, 2008 has been entered.

Response to Amendment

2.   **Claims 1**, **21**, **24-25**, **28-29**, and **31** have been amended.

        Claims 1-31 are pending in this instant Office action.

Response to Arguments

3.   Applicant's arguments with respect to the claims have been considered but are moot in view of the new ground(s) of rejection.

Claim Objections

4.   **Claim 25** is objected to because of the following informalities: wording error when citing "*A computer-readable record medium having stored thereon computer-executable instructions for causing a computer to perform <u>a method processing</u> a synonym ..*" Applicant is suggested to change from "... <u>a method processing</u> ..." to " ... <u>a method for processing</u> ..."

Claim Rejections - 35 USC § 112

5.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

6.   **Claims 1-31** are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Regarding **claim 1**, the claim recites a conditional statement "*if the predetermined synonym acceptance criterion is fulfilled*" on line 9, and the claimed language clearly states that the steps of "*associating the value of the data field ...*" and "*automatically updating a synonym set ...*" on lines 9-14 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on line 9 fails is not specified any where in the claimed language.

Regarding **claim 13**, the claim recites a conditional statement "*when the predetermined synonym acceptance criterion is fulfilled*" on lines 4-5, and the claimed language clearly states that the steps of "*the value of the data field is added to the synonym set ...*" can be only carried out if the "*when*" conditional statement above is true. What happens when the "*when*" condition on lines 4-5 fails is not specified any where in the claimed language.

Regarding **claim 14**, the claim recites a conditional statement "*if a predetermined discard criterion is fulfilled*" on lines 2-3, and the claimed language clearly states that the steps of "*determining the at least one synonym candidate is discarded*" can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on lines 2-3 fails is not specified any where in the claimed language.

Regarding **claim 21**, the claim recites a conditional statement "*if the value of the data field and a synonym candidate fulfill a predetermined synonym acceptance criterion taking into account writing variations*" on lines 9-10, and the claimed language clearly states that the step of "*automatically adding before searching...*" on lines 10-12 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on lines 9-10 fails is not specified any where in the claimed language.

Regarding **claim 24**, the claim recites a conditional statement "*if the predetermined synonym acceptance criterion is fulfilled*" on line 12, and the claimed language clearly states that the steps of "*associating the value of the data field ...*" and "*automatically updating a synonym set ...*" on lines 12-16 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on line 12 fails is not specified any where in the claimed language.

Regarding **claim 25**, the claim recites a conditional statement "*if the value of the data field and a synonym candidate fulfill a predetermined synonym acceptance criterion taking into account writing variations*" on lines 11-13, and the claimed language clearly states that the step of "*automatically adding before searching...*" on lines 13-15 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on lines11-13 fails is not specified any where in the claimed language.

Regarding **claim 26**, the claim recites a conditional statement "*if the predetermined synonym acceptance criterion is fulfilled*" on line 13, and the claimed

language clearly states that the steps of "*associate the value of the data field ...*" and "*automatically add the synonym candidate ...*" on lines 13-17 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on line 13 fails is not specified any where in the claimed language.

Regarding **claim 28**, the claim recites a conditional statement "*if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion taking into account writing variations*" on lines 15-17, and the claimed language clearly states that the step of "*associate the value of the data field and the synonym candidate as synonyms*" on lines 18-19 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on lines15-17 fails is not specified any where in the claimed language.

Regarding **claim 29**, the claim recites a conditional statement "*if the predetermined synonym acceptance criterion is fulfilled*" on lines 8-9, and the claimed language clearly states that the steps of "*associate the value of the data field ...*" and "*automatically add the synonym candidate ...*" on lines 9-13 can be only carried out if the "*if*" conditional statement above is true. What happens when the "*if*" condition on lines 8-9 fails is not specified any where in the claimed language.

Further, the claim recites an obscured statement of "*... provide an updated a synonym ...*" on line 12. Clarification is required.

Regarding **claim 31**, the claim recites a conditional statement "*if the predetermined synonym acceptance criterion is fulfilled*" on line 5, and the claimed language clearly states that the steps of "*associate the value of the data field ...*" and

"*automatically add to a synonym set ...*" on lines 5-9 can be only carried out if the "*if*"

conditional statement above is true. What happens when the "*if*" condition on line 5 fails

is not specified any where in the claimed language.

Claims 2-20, 22-23, 27, and 30 fail to resolve the deficiencies of claims 1, 21,

26, and 29 respectively since they only further limit the scope of claims 1, 21, 26, and

29. Therefore, claims 2-20, 22-23, 27, and 30 are also rejected under 35 U.S.C. 112,

2nd paragraph.

Claim Rejections - 35 USC § 101

7.   35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

8.   Claims 1-23, 26-28, and 31 are rejected under 35 U.S.C. 101 because the

claimed invention is directed to non-statutory subject matter.

Regarding claim 1, the claim recites a method for processing data records

without confining its method steps to a "particular machine" (e.g. a general-purpose

computer). Applicant is suggested to have at least a computer storage component in the

claimed method and explain how the at least storage component can be utilized with the

method steps.

Regarding claim 21, the claim recites a method for processing a synonym set

without confining its method steps to a "particular machine" (e.g. a general-purpose

computer). Applicant is suggested to have at least a computer storage component in the

claimed method and explain how the at least storage component can be utilized with the method steps.

Claims 2-20, and 22-23 fail to resolve the deficiencies of claims 1, and 21 respectively since they only further limit the scope of claims 1, and 21. Therefore, claims 2-20, and 22-23 are also rejected under 35 U.S.C. 101.

Regarding claims 26, 28, and 31, each claim recites a "*data processing system*" or a "*data processing apparatus*". However, it is not clear whether this claimed system/apparatus is equipped with at least a processor and a memory, wherein the memory stores thereon a computer program, and the at least one processor operates according to this computer program. Applicant is suggested to include at least a processor and a memory, and explain how they can be utilized in the claimed "*data processing system/apparatus*".

The claimed system/apparatus is directed to a software system itself, not a process occurring as a result of actually executing the software components, a machine programmed to operate in accordance with the software components, nor a manufacture structurally and functionally interconnected with the software components in a manner which enables the software components to carry out their functionalities. The claimed system/apparatus is also not a combination of chemical compounds to be a composition of matter. As such, it fails to fall within a statutory category. It is, at best, functional descriptive material *per se*.

Claim 27 fails to resolve the deficiencies of claim 26 since it only further limits the scope of claim 26. Therefore, claim 27 is also rejected under 35 U.S.C. 101.

The claims above lack the necessary physical articles or objects to constitute a machine or a manufacture within the meaning of 35 U.S.C. 101. They are clearly not a series of steps or acts to be a process nor are they a combination of chemical compounds to be a composition of matter. As such, they fail to fall within a statutory category. They are, at best, functional descriptive material *per se*.

Descriptive material can be characterized as either "functional descriptive material" or "nonfunctional descriptive material." Both types of "descriptive material" are nonstatutory when claimed as descriptive material *per se*, 33 F.3d at 1360, 31 USPQ2d at 1759. When functional descriptive material is recorded on some computer-readable medium, it becomes structurally and functionally interrelated to the medium and will be statutory in most cases since use of technology permits the function of the descriptive material to be realized. Compare *In re Lowry*, 32 F.3d 1579, 1583-84, 32 USPQ2d 1031, 1035 (Fed. Cir. 1994)

Merely claiming nonfunctional descriptive material, i.e., abstract ideas, stored on a computer-readable medium, in a computer, or on an electromagnetic carrier signal, does not make it statutory. See Diehr, 450 U.S. at 185-86, 209 USPQ at 8 (noting that the claims for an algorithm in *Benson* were unpatentable as abstract ideas because "[t]he sole practical application of the algorithm was in connection with the programming of a general purpose computer.")

Claim Rejections - 35 USC § 103

9.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10.   This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

11.   **Claims 1**-**6**, **9**, and **13**-**31** are rejected under 35 U.S.C. 103(a) as being unpatentable over Simske et al. (*Pub. No. US 2004/0064447, filed on September 27, 2002; hereinafter Simske*) in view of Toner et al. (*Pub. No. US 2004/0024760, filed on July 31, 2002; hereinafter Toner*).

Regarding **claim 1**, Simske clearly shows and discloses a method of processing a data record for finding a counterpart in a reference data set (*Figure 6*), the method comprising the steps of:

determining a value of a data field, the data field representing an identifier (*Suppose the user inputs the query "class list for Stanford"*, [0101]),

determining from a set of predetermined identifier values at least one synonym candidate for the value of the data field (*For the term "class", the following synonyms*

are identified by the synonymic search application: set, group, division, grade, rank,

category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8

candidate terms (including the word "class" itself) that may be used in searching for

"class". For the term "list", the following synonyms are identified by the synonymic

search application: catalog, inventory, register, record, roll, and directory. Thus, 6

synonyms are identified for the term "list", resulting in 7 candidate terms (including the

word "list" itself) that may be used in searching for "list", [0101]),

determining if a synonym candidate and the value of the data field fulfill a

predetermined synonym acceptance criterion (*supposing that no more than 25 queries*

are allowed (e.g., because of the user-tuned breadth of the synonymic search query to

be performed and/or because of the synonymic search application's implemented query

limits), the above-identified 56 queries need to be reduced to the 25 optimal queries to

be utilized, [0101]), and if the predetermined synonym acceptance criterion is fulfilled,

associating the value of the data field and the synonym candidate as synonyms (*the*

synonym database (i.e., the electronic thesaurus or other source from which synonyms

are determined) is structured such that the synonyms are rated for their "closeness in

meaning" or "proximity" to the original word. Such rating may be performed by the

electronic thesaurus, the synonymic search application, some other application, or any

combination thereof. For example, suppose such statistics are available for "class" and

"list", then the various synonyms for each of the terms may be weighted based on their

relative proximity to their respective base word (i.e., "class" or "list"), [0103]) and

automatically updating a synonym set associated with the value of the data field by

adding the synonym candidate to the synonym set without intervention of a user before

searching for a counterpart (*Once the weighting for each possible synonymic query is*

determined in block 605 of Figure 6 (e.g., by multiplying the assigned weight value for

each word of the query), the highest weighted "Q" queries to be included in the

constructed synonymic search query are determined in block 606. For instance, in the

above example, the highest weighted 25 synonymic queries (which includes the original

user-input query itself) are determined for inclusion in the constructed synonymic search

query, [0116]),

searching for the counterpart for the data record by comparing to entries of the

reference data set the value of the data field and/or the synonym set after the step of

determining if the predetermined synonym acceptance criterion is fulfilled (*the overall*

relevance of all the search results is determined by comparing its keywords to those in

the original, user-input query. The top result(s) of each individual query included in a

synonymic search query may be presented to a user, which may widen the breadth of

the search query--e.g., provides a trade-off between overall weight and weight within a

novel query, [0145]), wherein if the synonym set was updated, said comparison to the

synonym set comprises comparison to the updated synonym set (*a number "Q" of*

queries to be included in the synonymic search query is determined based at least in

part on the breadth desired for the synonymic search query. If the user tunes the

breadth of the synonymic search query to be very general, then the number "Q" may be

determined to be much larger (e.g., 25 or more), or the user may tune the breadth to

any other amount desired. Thus, the tuning of the breadth of the synonymic search

query in block 303 may dictate the total number of queries to be included in the

constructed synonymic search query, [0096]. It is well inherent that if the breath of the

query is greater than 26, then the top 26 queries are produced for searching for

counterparts).

Simske does not explicitly disclose the data field is from a record, and the

synonyms take into account writing variations.

However, Toner discloses the data field is from a record (*Users select whether*

they wish to use a single Suspect name at a time, manually entered by the keyboard, or

to use an existing file containing a list of Suspect names, [0094]), and synonyms take

into account writing variations (*The conversion process may be done by finding the*

nearest name in a list of pre-defined, standard names, or by using an algorithm and

techniques to do the conversion in real time into a form that caters for all possible

variations of spelling and splitting and concatenations. The results are given a

confidence level (from a string correlation function) presented as a percentage, [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of

the invention was made to incorporate the teachings of Toner with the teachings of

Simske for the purpose of resolving variations in spellings and representations for

names with foreign origins that may be spelt in any number of ways ([0015] of Toner).

Regarding **claim 2**, Simske further discloses the at least one synonym candidate

is determined using a candidate selection criterion depending at least on the value of

the data field and on a synonym candidate (*i.e., the electronic thesaurus or other source*

from which synonyms are determined) is structured such that the synonyms are rated

for their "closeness in meaning" or "proximity" to the original word. Such rating may be

performed by the electronic thesaurus, the synonymic search application, some other

application, or any combination thereof. For example, suppose such statistics are

available for "class" and "list", then the various synonyms for each of the terms may be

weighted based on their relative proximity to their respective base word (i.e., "class" or

"list").

Regarding **claim 3**, <u>Toner</u> further discloses the candidate selection criterion

takes into account how similar a synonym candidate and the value of the data field

sound (*The rules employed take into account the original sounds or pronunciations of*

the letters, eliminating double letters, and looking for special patterns, [0062]).

Regarding **claim 4**, <u>Toner</u> further discloses the candidate selection criterion

specifies that at least a predetermined part of the value of the data field is identical to a

predetermined part of a synonym candidate (*a probabilistic search algorithm is used*

that matches strings according to the length and number of string fragments shared by

the two strings, [0062]).

Regarding **claim 5**, <u>Simske</u> further discloses the candidate selection criterion

takes into account also a further data field of the data record, said further data field

representing a second identifier (*For the term "class", the following synonyms are*

identified by the synonymic search application: set, group, division, grade, rank,

category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8

candidate terms (including the word "class" itself) that may be used in searching for

"class". For the term "list", the following synonyms are identified by the synonymic

search application: catalog, inventory, register, record, roll, and directory. Thus, 6

synonyms are identified for the term "list", resulting in 7 candidate terms (including the

word "list" itself) that may be used in searching for "list", [0101]).

Regarding **claim 6**, <u>Simske</u> further discloses at least one quality parameter is

evaluated for a synonym candidate, the synonym acceptance criterion taking into

account the at least one quality parameter (*the synonymic search application may be*

operable to autonomously weight the synonymic queries in the manner described more

filly below in conjunction with Figure 6 such that the optimal synonymic queries are

more heavily weighted, [0078]).

Regarding **claim 9**, <u>Toner</u> further discloses the proportion of identical characters

takes into account the order of the characters (*a probabilistic search algorithm is used*

that matches strings according to the length and number of string fragments shared by

the two strings, [0062]).

Regarding **claim 13**, <u>Simske</u> further discloses a method, wherein the search for

the counterpart involves comparison of the value of the data field to a synonym set

relating to the identifier, members of said synonym set referring to respective

predetermined identifier values, and when the predetermined synonym acceptance

criterion is fulfilled, the value of the data field is added to the synonym set as a member

referring to the synonym associated with the value of the data field before the search for

the counterpart (*For the term "class", the following synonyms are identified by the synonymic search application: set, group, division, grade, rank, category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8 candidate terms (including the word "class" itself) that may be used in searching for "class". For the term "list", the following synonyms are identified by the synonymic search application: catalog, inventory, register, record, roll, and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate terms (including the word "list" itself) that may be used in searching for "list"*, [0101]).

Regarding **claim 14**, Simske further discloses wherein determining the at least one synonym candidate is discarded, if a predetermined discard criterion is fulfilled (*supposing that no more than 25 queries are allowed (e.g., because of the user-tuned breadth of the synonymic search query to be performed and/or because of the synonymic search application's implemented query limits), the above-identified 56 queries need to be reduced to the 25 optimal queries to be utilized*, [0101]).

Regarding **claim 15**, Toner further discloses the predetermined discard criterion specifies that the value of the data field is identical to one of the predetermined identifier values (*see the example of permutation process of name components in Figure 6*).

Regarding **claim 16**, Simske further discloses the search for the counterpart involves the synonym set and the predetermined discard criterion specifies that the value of the data field is at least one of the following: one of the predetermined identifier values, and a member of the synonym set (*the overall relevance of all the search results*

is determined by comparing its keywords to those in the original, user-input query. The

top result(s) of each individual query included in a synonymic search query may be

presented to a user, which may widen the breadth of the search query--e.g., provides a

trade-off between overall weight and weight within a novel query, [0145]).

Regarding **claim 17**, <u>Simske</u> further discloses the predetermined discard criterion

takes into account a value of a second data field in the data record (*Once the weighting*

for each possible synonymic query is determined in block 605 of Figure 6 (e.g., by

multiplying the assigned weight value for each word of the query), the highest weighted

"Q" queries to be included in the constructed synonymic search query are determined in

block 606. For instance, in the above example, the highest weighted 25 synonymic

queries (which includes the original user-input query itself) are determined for inclusion

in the constructed synonymic search query, [0116]).

Regarding **claim 18**, <u>Simske</u> further discloses information indicating the at least

one synonym associated with the value of the data field is added to the data record

(*Once the weighting for each possible synonymic query is determined in block 605 of*

Figure 6 (e.g., by multiplying the assigned weight value for each word of the query), the

highest weighted "Q" queries to be included in the constructed synonymic search query

are determined in block 606. For instance, in the above example, the highest weighted

25 synonymic queries (which includes the original user-input query itself) are

determined for inclusion in the constructed synonymic search query, [0116]).

Regarding **claim 19**, <u>Simske</u> further discloses a method, wherein a copy of the

data record is made for each synonym associated with the value of the data field (*each*

of the plurality of search engines may be utilized to perform the constructed synonymic

search query, [0079]).

Regarding **claim 20**, <u>Toner</u> further discloses a method, wherein the identifier

relates to a name of one of the following: a geographical entity, a person and an

organisation (*compare the search query with the Name Database*, [0030]).

Regarding **claim 21**, <u>Simske</u> clearly shows and discloses a method of processing

a synonym set for searching counterparts in a reference data set for data records, a

data field representing an identifier, members of the synonym set being first identifier

values and referring to respective second identifier values, the second identifier values

being predetermined identifier values , and said searching for a counterpart involving

comparison of a value of the data field to the synonym set (*Figure 6*), the method

comprising the steps of, determining among the predetermined identifier values at least

one synonym candidate relating to the value of the data field (*For the term "class", the*

following synonyms are identified by the synonymic search application: set, group,

division, grade, rank, category, and order. Thus, 7 synonyms are identified for the term

"class", resulting in 8 candidate terms (including the word "class" itself) that may be

used in searching for "class". For the term "list", the following synonyms are identified by

the synonymic search application: catalog, inventory, register, record, roll, and directory.

Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate terms

(including the word "list" itself) that may be used in searching for "list", [0101]), if the

value of the data field and a synonym candidate fulfill a predetermined synonym

acceptance criterion taking into account writing variations (*supposing that no more than*

25 queries are allowed (e.g., because of the user-tuned breadth of the synonymic

search query to be performed and/or because of the synonymic search application's

implemented query limits), the above-identified 56 queries need to be reduced to the 25

optimal queries to be utilized, [0101]), automatically adding before searching for a

counterpart for a data record the value of the data field to the synonym set as a member

referring to the synonym candidate without intervention of a user (*the synonym*

database (i.e., the electronic thesaurus or other source from which synonyms are

determined) is structured such that the synonyms are rated for their "closeness in

meaning" or "proximity" to the original word. Such rating may be performed by the

electronic thesaurus, the synonymic search application, some other application, or any

combination thereof. For example, suppose such statistics are available for "class" and

"list", then the various synonyms for each of the terms may be weighted based on their

relative proximity to their respective base word (i.e., "class" or "list"), [0103]).

    Simske does not explicitly disclose the data field is from a record, and the

synonyms take into account writing variations.

    However, Toner discloses the data field is from a record (*Users select whether*

they wish to use a single Suspect name at a time, manually entered by the keyboard, or

to use an existing file containing a list of Suspect names, [0094]), and synonyms take

into account writing variations (*The conversion process may be done by finding the*

nearest name in a list of pre-defined, standard names, or by using an algorithm and

techniques to do the conversion in real time into a form that caters for all possible

variations of spelling and splitting and concatenations. The results are given a

confidence level (from a string correlation function) presented as a percentage, [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of

the invention was made to incorporate the teachings of <u>Toner</u> with the teachings of

<u>Simske</u> for the purpose of resolving variations in spellings and representations for

names with foreign origins that may be spelt in any number of ways ([0015] of <u>Toner</u>).

Regarding **claim 22**, <u>Simske</u> further discloses a method, wherein the synonym

set is empty before adding the value of the data field to the synonym set (*Once the*

weighting for each possible synonymic query is determined in block 605 of Figure 6

(e.g., by multiplying the assigned weight value for each word of the query), the highest

weighted "Q" queries to be included in the constructed synonymic search query are

determined in block 606. For instance, in the above example, the highest weighted 25

synonymic queries (which includes the original user-input query itself) are determined

for inclusion in the constructed synonymic search query, [0116]).

Regarding **claim 23**, <u>Simske</u> further discloses a method, wherein the synonym

set contains at least one member before adding the value of the data field to the

synonym set (*a number "Q" of queries to be included in the synonymic search query is*

determined based at least in part on the breadth desired for the synonymic search

query. If the user tunes the breadth of the synonymic search query to be very general,

then the number "Q" may be determined to be much larger (e.g., 25 or more), or the

user may tune the breadth to any other amount desired. Thus, the tuning of the breadth

of the synonymic search query in block 303 may dictate the total number of queries to

be included in the constructed synonymic search query, [0096]. *It is well inherent that if*

the breath of the query is greater than 26, then the top 26 queries are produced for

searching for counterparts).

Regarding **claim 24**, <u>Simske</u> further discloses a computer program embodied in

a computer-readable record medium, the computer-readable record medium including

program instructions for causing a computer to perform the method for processing a

data record for finding a counterpart in a reference data set ([0013]), the method

comprising the steps of:

determining a value of a data field, the data field representing an identifier

(*Suppose the user inputs the query "class list for Stanford",* [0101]),

determining from a set of predetermined identifier values at least one synonym

candidate for the value of the data field (*For the term "class", the following synonyms*

are identified by the synonymic search application: set, group, division, grade, rank,

category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8

candidate terms (including the word "class" itself) that may be used in searching for

"class". For the term "list", the following synonyms are identified by the synonymic

search application: catalog, inventory, register, record, roll, and directory. Thus, 6

synonyms are identified for the term "list", resulting in 7 candidate terms (including the

word "list" itself) that may be used in searching for "list", [0101]),

determining if a synonym candidate and the value of the data field fulfill a

predetermined synonym acceptance criterion (*supposing that no more than 25 queries*

are allowed (e.g., because of the user-tuned breadth of the synonymic search query to

be performed and/or because of the synonymic search application's implemented query

limits), the above-identified 56 queries need to be reduced to the 25 optimal queries to

be utilized, [0101]), and if the predetermined synonym acceptance criterion is fulfilled,

associating the value of the data field and the synonym candidate as synonyms (*the*

synonym database (i.e., the electronic thesaurus or other source from which synonyms

are determined) is structured such that the synonyms are rated for their "closeness in

meaning" or "proximity" to the original word. Such rating may be performed by the

electronic thesaurus, the synonymic search application, some other application, or any

combination thereof. For example, suppose such statistics are available for "class" and

"list", then the various synonyms for each of the terms may be weighted based on their

relative proximity to their respective base word (i.e., "class" or "list"), [0103]) and

automatically updating a synonym set associated with the value of the data field by

adding the synonym candidate to the synonym set without intervention of a user before

searching for a counterpart (*Once the weighting for each possible synonymic query is*

determined in block 605 of Figure 6 (e.g., by multiplying the assigned weight value for

each word of the query), the highest weighted "Q" queries to be included in the

constructed synonymic search query are determined in block 606. For instance, in the

above example, the highest weighted 25 synonymic queries (which includes the original

user-input query itself) are determined for inclusion in the constructed synonymic search query, [0116]),

searching for the counterpart for the data record by comparing to entries of the reference data set the value of the data field and/or the synonym set after the step of determining if the predetermined synonym acceptance criterion is fulfilled (*the overall relevance of all the search results is determined by comparing its keywords to those in the original, user-input query. The top result(s) of each individual query included in a synonymic search query may be presented to a user, which may widen the breadth of the search query--e.g., provides a trade-off between overall weight and weight within a novel query*, [0145]), wherein if the synonym set was updated, said comparison to the synonym set comprises comparison to the updated synonym set (*a number "Q" of queries to be included in the synonymic search query is determined based at least in part on the breadth desired for the synonymic search query. If the user tunes the breadth of the synonymic search query to be very general, then the number "Q" may be determined to be much larger (e.g., 25 or more), or the user may tune the breadth to any other amount desired. Thus, the tuning of the breadth of the synonymic search query in block 303 may dictate the total number of queries to be included in the constructed synonymic search query*, [0096]. *It is well inherent that if the breath of the query is greater than 26, then the top 26 queries are produced for searching for counterparts*).

Simske does not explicitly disclose the data field is from a record, and the synonyms take into account writing variations.

However, <u>Toner</u> discloses the data field is from a record (*Users select whether they wish to use a single Suspect name at a time, manually entered by the keyboard, or to use an existing file containing a list of Suspect names,* [0094]), and synonyms take into account writing variations (*The conversion process may be done by finding the nearest name in a list of pre-defined, standard names, or by using an algorithm and techniques to do the conversion in real time into a form that caters for all possible variations of spelling and splitting and concatenations. The results are given a confidence level (from a string correlation function) presented as a percentage,* [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of the invention was made to incorporate the teachings of <u>Toner</u> with the teachings of <u>Simske</u> for the purpose of resolving variations in spellings and representations for names with foreign origins that may be spelt in any number of ways ([0015] of <u>Toner</u>).

Regarding **claim 25**, <u>Simske</u> further discloses a computer-readable record medium having stored thereon computer-executable instructions for causing a computer to perform a method processing a synonym set for searching counterparts in a reference data set for data records, a data field representing an identifier, members of the synonym set being first identifier values and referring to respective second identifier values, the second identifier values being predetermined identifier values, and said searching for a counterpart involving comparison of a value of the data field to the synonym set ([0013]), the method comprising the steps of determining among the predetermined identifier values at least one synonym candidate relating to the value of the data field (*For the term "class", the following synonyms are identified by the*

synonymic search application: set, group, division, grade, rank, category, and order.

Thus, 7 synonyms are identified for the term "class", resulting in 8 candidate terms

(including the word "class" itself) that may be used in searching for "class". For the term

"list", the following synonyms are identified by the synonymic search application:

catalog, inventory, register, record, roll, and directory. Thus, 6 synonyms are identified

for the term "list", resulting in 7 candidate terms (including the word "list" itself) that may

be used in searching for "list", [0101]), and, if the value of the data field and a synonym

candidate fulfill a predetermined synonym acceptance criterion (*supposing that no more*

than 25 queries are allowed (e.g., because of the user-tuned breadth of the synonymic

search query to be performed and/or because of the synonymic search application's

implemented query limits), the above-identified 56 queries need to be reduced to the 25

optimal queries to be utilized, [0101]), automatically adding before searching for a

counterpart for a data record the value of the data field to the synonym set as a member

referring to the synonym candidate without intervention of a user (*the synonym*

database (i.e., the electronic thesaurus or other source from which synonyms are

determined) is structured such that the synonyms are rated for their "closeness in

meaning" or "proximity" to the original word. Such rating may be performed by the

electronic thesaurus, the synonymic search application, some other application, or any

combination thereof. For example, suppose such statistics are available for "class" and

"list", then the various synonyms for each of the terms may be weighted based on their

relative proximity to their respective base word (i.e., "class" or "list"), [0103]).

Simske does not explicitly disclose the data field is from a record, and the synonyms take into account writing variations.

However, Toner discloses the data field is from a record (*Users select whether they wish to use a single Suspect name at a time, manually entered by the keyboard, or to use an existing file containing a list of Suspect names*, [0094]), and synonyms take into account writing variations (*The conversion process may be done by finding the nearest name in a list of pre-defined, standard names, or by using an algorithm and techniques to do the conversion in real time into a form that caters for all possible variations of spelling and splitting and concatenations. The results are given a confidence level (from a string correlation function) presented as a percentage*, [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of the invention was made to incorporate the teachings of Toner with the teachings of Simske for the purpose of resolving variations in spellings and representations for names with foreign origins that may be spelt in any number of ways ([0015] of Toner).

Regarding **claim 26**, Simske clearly shows and discloses a data processing system for processing data records for finding counterparts in a reference data set ([0014]), the system comprising:

Simske discloses:

means for storing the reference data set (*Search engines enable a user to input a search query thereto and retrieve from the corpus of information (e.g., a local*

database and/or client-server network) information containing the user-specified search query terms, [0029]),

means for storing predetermined identifier values for an identifier (*For the term "class", the following synonyms are identified by the synonymic search application: set, group, division, grade, rank, category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8 candidate terms (including the word "class" itself) that may be used in searching for "class". For the term "list", the following synonyms are identified by the synonymic search application: catalog, inventory, register, record, roll, and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate terms (including the word "list" itself) that may be used in searching for "list"*, [0101]),

means for determining values of a data filed, the data field representing the identifier (*Suppose the user inputs the query "class list for Stanford"*, [0101]),

means for associating values of the field and respective predetermined identifier values as synonyms, said means configured to determine from the predetermined identifier values at least one synonym candidate for a value of the data field (*For the term "class", the following synonyms are identified by the synonymic search application: set, group, division, grade, rank, category, and order. Thus, 7 synonyms are identified for the term "class", resulting in 8 candidate terms (including the word "class" itself) that may be used in searching for "class". For the term "list", the following synonyms are identified by the synonymic search application: catalog, inventory, register, record, roll, and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate*

terms (including the word "list" itself) that may be used in searching for "list", [0101]), to
determine if a synonym candidate and the value of the data field fulfill a predetermined
synonym acceptance criterion (*supposing that no more than 25 queries are allowed*
(e.g., because of the user-tuned breadth of the synonymic search query to be
performed and/or because of the synonymic search application's implemented query
limits), the above-identified 56 queries need to be reduced to the 25 optimal queries to
be utilized, [0101]), and if the predetermined synonym acceptance synonym is fulfilled,
to associate the value of the data field and the synonym candidate as synonyms (*i.e.,*
the electronic thesaurus or other source from which synonyms are determined) is
structured such that the synonyms are rated for their "closeness in meaning" or
"proximity" to the original word. Such rating may be performed by the electronic
thesaurus, the synonymic search application, some other application, or any
combination thereof. For example, suppose such statistics are available for "class" and
"list", then the various synonyms for each of the terms may be weighted based on their
relative proximity to their respective base word (i.e., "class" or "list"), [0103]) and to
automatically add the synonym candidate to a synonym set associated with the value of
the data field without intervention of a user before searching for a counterpart to provide
an updated synonym set (*Once the weighting for each possible synonymic query is*
determined in block 605 of Figure 6 (e.g., by multiplying the assigned weight value for
each word of the query), the highest weighted "Q" queries to be included in the
constructed synonymic search query are determined in block 606. For instance, in the
above example, the highest weighted 25 synonymic queries (which includes the original

user-input query itself) are determined for inclusion in the constructed synonymic search

query, [0116]), and

means for searching counterparts in the reference data set for the data records

by comparing to entries of the reference data set values of data fields and/or said

updated synonyms set (*the overall relevance of all the search results is determined by*

comparing its keywords to those in the original, user-input query. The top result(s) of

each individual query included in a synonymic search query may be presented to a

user, which may widen the breadth of the search query--e.g., provides a trade-off

between overall weight and weight within a novel query, [0145]).

Simske does not disclose receive data records, and determining the data field

representing the identifier in the data records.

However, Toner discloses:

means for receiving data records (*Users select whether they wish to use a single*

Suspect name at a time, manually entered by the keyboard, or to use an existing file

containing a list of Suspect names, [0094]),

means for determining in the data records values of a data field, the data field

representing the identifier (*Figure 3 shows an exemplary parsed name where in the*

inputted suspect name "Robert James Smith" is divided into three name components

'Robert', 'James', and 'Smith', [0139]-[0140]),

It would have been obvious to an ordinary person skilled in the art at the time of

the invention was made to incorporate the teachings of <u>Toner</u> with the teachings of

<u>Simske</u> for the purpose of resolving variations in spellings and representations for

names with foreign origins that may be spelt in any number of ways ([0015] of <u>Toner</u>).

Regarding **claim 27**, <u>Simske</u> further discloses:

means for storing a synonym set, members of said synonym set referring to

respective predetermined identifier values (*a thesaurus compiles many words in the*

English language and identifies synonyms that may be used in place of each word,

[0030]),

wherein the means for associating values of the data field and respective

predetermined identifier values as synonyms are configured to add to the synonym set

a member referring to the synonym associated with the value of the data field before

activation of the means for searching counterparts (*the overall relevance of all the*

search results is determined by comparing its keywords to those in the original, user-

input query. The top result(s) of each individual query included in a synonymic search

query may be presented to a user, which may widen the breadth of the search query--

e.g., provides a trade-off between overall weight and weight within a novel query,

[0145]).

Regarding **claim 28**, <u>Simske</u> clearly shows and discloses a data processing

system for processing a synonym set for searching counterparts in a reference data set

for data records, a data field representing an identifier, members of the synonym set

being first identifier values and referring to respective second identifier values, said

second identifier values being predetermined identifier values, and said searching

involving comparing a value of the data field to the synonym set ([0014]), the system

comprising :

means for storing the synonym set (*a thesaurus compiles many words in the*

English language and identifies synonyms that may be used in place of each word,

[0030]),

means for storing predetermined identifier values for the identifier (*For the term*

"class", the following synonyms are identified by the synonymic search application: set,

group, division, grade, rank, category, and order. Thus, 7 synonyms are identified for

the term "class", resulting in 8 candidate terms (including the word "class" itself) that

may be used in searching for "class". For the term "list", the following synonyms are

identified by the synonymic search application: catalog, inventory, register, record, roll,

and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate

terms (including the word "list" itself) that may be used in searching for "list", [0101]),

means automatically for adding to the synonym set a value of the data field and

respective predetermined identifier values associated as synonyms without intervention

of a user before searching for counterparts in the reference data set (*For the term*

"class", the following synonyms are identified by the synonymic search application: set,

group, division, grade, rank, category, and order. Thus, 7 synonyms are identified for

the term "class", resulting in 8 candidate terms (including the word "class" itself) that

may be used in searching for "class". For the term "list", the following synonyms are

identified by the synonymic search application: catalog, inventory, register, record, roll,

and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate

terms (including the word "list" itself) that may be used in searching for "list", [0101]),

said means configured to determine from the predetermined identifier values at

least one synonym candidate for a value of the data field (*For the term "class", the*

following synonyms are identified by the synonymic search application: set, group,

division, grade, rank, category, and order. Thus, 7 synonyms are identified for the term

"class", resulting in 8 candidate terms (including the word "class" itself) that may be

used in searching for "class". For the term "list", the following synonyms are identified by

the synonymic search application: catalog, inventory, register, record, roll, and directory.

Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate terms

(including the word "list" itself) that may be used in searching for "list", [0101]), to

determine if a synonym candidate and the value of the data field fulfill a predetermined

synonym acceptance criterion (*supposing that no more than 25 queries are allowed*

(e.g., because of the user-tuned breadth of the synonymic search query to be

performed and/or because of the synonymic search application's implemented query

limits), the above-identified 56 queries need to be reduced to the 25 optimal queries to

be utilized, [0101]), and if the predetermined synonym acceptance criterion is fulfilled, to

associate the value of the data field and the synonym candidates as synonyms (*For the*

term "class", the following synonyms are identified by the synonymic search application:

set, group, division, grade, rank, category, and order. Thus, 7 synonyms are identified

for the term "class", resulting in 8 candidate terms (including the word "class" itself) that may be used in searching for "class". For the term "list", the following synonyms are identified by the synonymic search application: catalog, inventory, register, record, roll, and directory. Thus, 6 synonyms are identified for the term "list", resulting in 7 candidate terms (including the word "list" itself) that may be used in searching for "list", [0101]).

Simske does not disclose receiving data records, and determining the data field representing the identifier in the data records.

However, Toner discloses:

receiving data records (*Users select whether they wish to use a single Suspect name at a time, manually entered by the keyboard, or to use an existing file containing a list of Suspect names*, [0094]),

determining in the data records values of a data field, the data field representing the identifier (*Figure 3 shows an exemplary parsed name where in the inputted suspect name "Robert James Smith" is divided into three name components 'Robert', 'James', and 'Smith'*, [0139]-[0140]),

It would have been obvious to an ordinary person skilled in the art at the time of the invention was made to incorporate the teachings of Toner with the teachings of Simske for the purpose of resolving variations in spellings and representations for names with foreign origins that may be spelt in any number of ways ([0015] of Toner).

Regarding **claim 29**, <u>Simske</u> clearly shows and discloses a data processing

apparatus, comprising at least one processor configured to process data records for

finding counterparts in a reference data set ([0014]), to determine values of a data field,

the data field representing an identifier (*Suppose the user inputs the query "class list for*

Stanford", [0101]), to associate values of the data field and respective predetermined

identifier values as synonyms (*Name substitution matching, where component words of*

the Suspect name are checked against a synonym table and are replaced with their

respective synonyms, [0088]), to determine if a synonym candidate and the value of the

data field fulfill a predetermined synonym acceptance criterion (*supposing that no more*

than 25 queries are allowed (e.g., because of the user-tuned breadth of the synonymic

search query to be performed and/or because of the synonymic search application's

implemented query limits), the above-identified 56 queries need to be reduced to the 25

optimal queries to be utilized, [0101]), and if the predetermined synonym acceptance

criterion is fulfilled, to associate the value of the data field and the synonym candidate

as synonyms (*the synonym database (i.e., the electronic thesaurus or other source from*

which synonyms are determined) is structured such that the synonyms are rated for

their "closeness in meaning" or "proximity" to the original word. Such rating may be

performed by the electronic thesaurus, the synonymic search application, some other

application, or any combination thereof. For example, suppose such statistics are

available for "class" and "list", then the various synonyms for each of the terms may be

weighted based on their relative proximity to their respective base word (i.e., "class" or

"list"), [0103]), and to automatically add the synonym candidate to a synonym set

associated with the value of the data field to provide an updated a synonym without

intervention of a user before searching for counterparts (*Once the weighting for each*

possible synonymic query is determined in block 605 of Figure 6 (e.g., by multiplying the

assigned weight value for each word of the query), the highest weighted "Q" queries to

be included in the constructed synonymic search query are determined in block 606.

For instance, in the above example, the highest weighted 25 synonymic queries (which

includes the original user-input query itself) are determined for inclusion in the

constructed synonymic search query, [0116]), to store the updated synonym set, and to

search the counterparts in the reference data set for the data records by comparing the

data records to entries of the reference data set values of data fields and/or said

updated synonym set (*the overall relevance of all the search results is determined by*

comparing its keywords to those in the original, user-input query. The top result(s) of

each individual query included in a synonymic search query may be presented to a

user, which may widen the breadth of the search query--e.g., provides a trade-off

between overall weight and weight within a novel query, [0145]).

Simske does not explicitly disclose the data field is from a record, and the

synonyms take into account writing variations.

However, Toner discloses the data field is from a record (*Users select whether*

they wish to use a single Suspect name at a time, manually entered by the keyboard, or

to use an existing file containing a list of Suspect names, [0094]), and synonyms take

into account writing variations (*The conversion process may be done by finding the*

nearest name in a list of pre-defined, standard names, or by using an algorithm and

techniques to do the conversion in real time into a form that caters for all possible

variations of spelling and splitting and concatenations. The results are given a

confidence level (from a string correlation function) presented as a percentage, [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of

the invention was made to incorporate the teachings of <u>Toner</u> with the teachings of

<u>Simske</u> for the purpose of resolving variations in spellings and representations for

names with foreign origins that may be spelt in any number of ways ([0015] of <u>Toner</u>).

Regarding **claim 30**, <u>Simske</u> further discloses members of said synonym set

referring to respective predetermined identifier values, and wherein the at least one

processor is configured to add to the synonym set stored in the at least one memory a

member referring to the synonym associated with the value of the data field before

activation of the search for counterparts (*Once the weighting for each possible*

synonymic query is determined in block 605 of Figure 6 (e.g., by multiplying the

assigned weight value for each word of the query), the highest weighted "Q" queries to

be included in the constructed synonymic search query are determined in block 606.

For instance, in the above example, the highest weighted 25 synonymic queries (which

includes the original user-input query itself) are determined for inclusion in the

constructed synonymic search query, [0116]).

Regarding **claim 31**, <u>Simske</u> clearly discloses a data processing apparatus

([0014]) configured to determine if a synonym candidate and the value of the data field

fulfill a predetermined synonym acceptance criterion (*supposing that no more than 25*

queries are allowed (e.g., because of the user-tuned breadth of the synonymic search

query to be performed and/or because of the synonymic search application's

implemented query limits), the above-identified 56 queries need to be reduced to the 25

optimal queries to be utilized, [0101]), and if the predetermined synonym acceptance

criterion is fulfilled, to associate the value of the data field and the synonym candidate

as synonyms (*the synonym database (i.e., the electronic thesaurus or other source from*

which synonyms are determined) is structured such that the synonyms are rated for

their "closeness in meaning" or "proximity" to the original word. Such rating may be

performed by the electronic thesaurus, the synonymic search application, some other

application, or any combination thereof. For example, suppose such statistics are

available for "class" and "list", then the various synonyms for each of the terms may be

weighted based on their relative proximity to their respective base word (i.e., "class" or

"list"), [0103]), and to automatically add to a synonym set stored in a memory a value of

the data field and respective predetermined identifier values associated as synonyms

without intervention of a user to update the synonym set before input into a searching

system configured to search for counterparts in the reference data set by comparing a

value of the data field to the updated synonym set (*Once the weighting for each*

possible synonymic query is determined in block 605 of Figure 6 (e.g., by multiplying the

assigned weight value for each word of the query), the highest weighted "Q" queries to

be included in the constructed synonymic search query are determined in block 606.

For instance, in the above example, the highest weighted 25 synonymic queries (which

includes the original user-input query itself) are determined for inclusion in the

constructed synonymic search query, [0116]).

Simske does not explicitly disclose the synonyms take into account writing

variations.

However, Toner discloses the synonyms take into account writing variations (*The*

conversion process may be done by finding the nearest name in a list of pre-defined,

standard names, or by using an algorithm and techniques to do the conversion in real

time into a form that caters for all possible variations of spelling and splitting and

concatenations. The results are given a confidence level (from a string correlation

function) presented as a percentage, [0061]).

It would have been obvious to an ordinary person skilled in the art at the time of

the invention was made to incorporate the teachings of Toner with the teachings of

Simske for the purpose of resolving variations in spellings and representations for

names with foreign origins that may be spelt in any number of ways ([0015] of Toner).

12.   **Claims 7**-8 is rejected under 35 U.S.C. 103(a) as being unpatentable over

Simske et al. (*Pub. No. US 2004/0064447, filed on September 27, 2002; hereinafter*

Simske) in view of Toner et al. (*Pub. No. US 2004/0024760, filed on July 31, 2002;*

hereinafter Toner), and further in view of Bellany et al. (*Pub. No. US 2002/0078024,*

filed on October 12, 2001; hereinafter Bellany).

Regarding **claim 7**, Simske, as modified by Toner, does not teach at least one

quality parameter takes into account at least one of the following quantities: a number of

changes required for converting the value of the data field to be identical to a synonym

candidate; a proportion of identical characters in the value of the data field and in a

synonym candidate; and a difference between the length of the value of the data field

and the length of a synonym candidate.

However, Bellany discloses at least one quality parameter takes into account at

least one of the following quantities: a number of changes required for converting the

value of the data field to be identical to a synonym candidate (*on not finding any entries*

in the dictionary identical to the input data, the processor may then search allowing for

one error at first, and if that search fails, performing a further search, allowing for two

errors, and so on. A single error may be counted if the search term and the dictionary

entry differ by one character being deleted, added or replaced with a different character.

The quality of correspondence between two terms may be judged by calculating the "

Levenshtein" distance between the two strings, [0044]).

It would have been obvious to a person with ordinary skills in the art at the time

of the invention to incorporate the teachings of Bellany with the teachings of Simske, as

modified by Toner, for the purpose of retrieving a desired postal address from a plurality

of postal addresses by searching a dictionary for entries in the dictionary corresponding

to the searched terms ([Abstract] of Bellany).

Regarding **claim 8**, Bellany further discloses the number of changes required for

converting the value of the data field to be identical to a synonym candidate is

calculated using the Levenshtein distance (*on not finding any entries in the dictionary*

identical to the input data, the processor may then search allowing for one error at first,

and if that search fails, performing a further search, allowing for two errors, and so on. A

single error may be counted if the search term and the dictionary entry differ by one

character being deleted, added or replaced with a different character. The quality of

correspondence between two terms may be judged by calculating the " Levenshtein"

distance between the two strings, [0044]).

13.   **Claims 10-12** are rejected under 35 U.S.C. 103(a) as being unpatentable over

Simske et al. (*Pub. No. US 2004/0064447, filed on September 27, 2002; hereinafter*

Simske) in view of Toner et al. (*Pub. No. US 2004/0024760, filed on July 31, 2002;*

hereinafter Toner), and further in view of Murakami et al. (*Pub. No. US 2004/0181759,*

filed on July, 19, 2002; hereinafter Murakami).

Regarding **claim 10**, Simske, as modified by Toner, does not disclose a first

quality parameter is evaluated for each synonym candidate and at least a second

quality parameter is evaluated at least for the synonym candidate(s) having the best first

quality parameter.

However, Murakami further discloses a first quality parameter is evaluated for

each synonym candidate and at least a second quality parameter is evaluated at least

for the synonym candidate(s) having the best first quality parameter (*generating a first*

set of candidate synonyms for the object word, based on whole of the document data

and generating at least one second set of candidate synonyms for the object word,

based on at least one part of the document data and narrowing the candidate synonyms

contained in the first set using the candidate synonyms contained in the second set,

[0016]).

It would have been obvious to a person with ordinary skills in the art at the time of the

invention to incorporate the teachings of <u>Murakami</u> with the teachings of <u>Simske</u>, as modified by

<u>Toner</u>, for the purpose of generating the candidate synonyms more efficiently by handling all

abbreviations and peculiar terms including misspelled or misconverted words ([0015] of

<u>Murakami</u>).

Regarding **claim 11**, <u>Murakami</u> further discloses the synonym acceptance criterion

requires that there is only one synonym candidate having the best at least one quality parameter

(*Table 1 shows that the firstly ranked candidate "batt" has the highest degree of relatedness*,

[0055]).

Regarding **claim 12**, <u>Murakami</u> further discloses a method, wherein at least two quality

parameters are evaluated for each synonym candidate and the synonym candidate acceptance

criterion specifies a threshold for one of the at least two quality parameters, the threshold being

dependent on a further one of the at least two quality parameters (*generating a first set of*

candidate synonyms for the object word, based on whole of the document data and generating

at least one second set of candidate synonyms for the object word, based on at least one part of

the document data and narrowing the candidate synonyms contained in the first set using the

candidate synonyms contained in the second set, [0016] *Candidate synonyms of the first set can*

be narrowed with the candidate synonyms of the second sets similarly to the aforementioned

method. In this case, the candidates which are ranked in places equal to or higher than a

threshold value place in the second sets are evaluated to be "absolute." The candidate

synonyms evaluated to be "absolute" are almost regarded as synonyms, [0021]).

Conclusion

14.   These following prior arts made of record and not relied upon are considered

pertinent to Applicant's disclosure:

Burdick et al. (*Pub. No. US 2004/0107205*) teaches Boolean rule-based system

for clustering similar records.

Lach et al. (*Pub. No. US 2004/0088157*) teaches method for

characterizing/classifying a document.

Kato et al. (*Pat. No. 5,519,857*) teaches hierarchical pre-search type text search

method and apparatus and magnetic disk unit used in the apparatus.

The Examiner requests, in response to this Office action, support(s) must be

shown for language added to any original claims on amendment and any new claims.

That is, indicate support for newly added claim language by specifically pointing to

page(s) and line no(s) in the specification and/or drawing figure(s). This will assist the

Examiner in prosecuting the application.

When responding to this office action, Applicant is advised to clearly point out the

patentable novelty which he or she thinks the claims present, in view of the state of the

art disclosed by the references cited or the objections made. He or she must also show

how the amendments avoid such references or objections See 37 CFR 1.111(c).

Contact Information

15.   Any inquiry concerning this communication or earlier communications from the

Examiner should be directed to Son T. Hoang whose telephone number is (571) 270-

1752. The Examiner can normally be reached on Monday - Friday (7:30 AM – 5:00 PM).

If attempts to reach the Examiner by telephone are unsuccessful, the Examiner's

supervisor, Christian Chace can be reached on (571) 272-4190. The fax phone number

for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for published

applications may be obtained from either Private PAIR or Public PAIR. Status

information for unpublished applications is available through Private PAIR only. For more

information about the PAIR system, see http://pair-direct.uspto.gov. Should you have

questions on access to the Private PAIR system, contact the Electronic Business Center

(EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer

Service Representative or access to the automated information system, call 800-786-

9199 (IN USA OR CANADA) or 571-272-1000.

/Son T Hoang/
Examiner, Art Unit 2165
November 26, 2008

/S. P./
Primary Examiner, Art Unit 2164

/Christian P. Chace/
Supervisory Patent Examiner, Art Unit 2165